Name: Yutong He

Net ID: yhe29

Student ID: 29433162

E-MAIL: yhe29@u.rochester.edu

Course: CSC 240

DATE: Oct.24th , 2016

Project 1

MAIN CONTINENT IN THE FOLDER:

test\_short.txt

Short test dataset from textbook problem 6.6.

test.txt

Test dataset generated by my classmate Xuan Tang.

adult.txt

Required adult dataset from UCI database. (originally named adult.data)

OUTPUT.pdf

Output of the program for the adult dataset with minimum support of 0.6.

src (folder)

java source files:

Apriori.java

Class implementing Apriori algorithm which containing a main method

FPGrowth.java

Class implementing FP-Growth algorithm which containing a main method

FPTree.java

Class of FPTree object

HeaderNode.java

Class of header node object used to generate the header table in FP-Growth

HeaderComparator.java

Class of comparator used to sort the header table

AprioriImproved.java

Class implementing my improved Apriori algorithm which containing a main method

BRIEF DESCRIPTION

This program implements Apriori, FP-Growth, my improved Apriori algorithms.

Apriori and FP-Growth are generally based on the description and the pseudocode provided in the textbook.

For my improved algorithm, I used the hash table improvement and transaction scan reduction improvement strategies, for more details, please see my report and code.

HOW TO COMPILE AND RUN THE CODE

javac Apriori.java

java Apriori

OR

javac FPGrowth.java

java FPGrowth

OR

javac AprioriImproved.java

java AprioriImproved

CITATION:

Data Set:

Adult Data Set, <https://archive.ics.uci.edu/ml/datasets/Adult> , Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

TextBook:

Jiawei Han, Micheline, Kamber, Jian Pei, Data Mining: Concepts and Techniques, 3/E, Morgan Kaufmann, 2011

Open Source:

Monperrus, Martin. Java implementation of the Apriori algorithm for mining frequent itemsets. GitHub repository, <https://gist.github.com/monperrus/7157717>

Nobahar, Kamran. An implementation of FP-Growth algorithm in Java, GitHub repository, <https://github.com/goodinges/FP-Growth-Java> (referred for my insert method in FP-Growth)

[Generating all permutations of a given string](http://stackoverflow.com/questions/4240080/generating-all-permutations-of-a-given-string), Stack Overflow, <http://stackoverflow.com/questions/4240080/generating-all-permutations-of-a-given-string> (referred for my combinations generation method in FP-Growth)